Universality of citation distributions: Toward an objective measure of scientific impact

Filippo Radicchi*, Santo Fortunato*, and Claudio Castellanoh,k,1

*Complex Networks Lagrange Laboratory, Institute for Scientific Interchanged Foundation, 10133 Torino, Italy; and hCentre for Statistical Mechanics and Complexity, National Institute for the Physics of Matter-Consiglio Nazionale delle Ricerche, and Dipartimento di Fisica, “Sapienza” Università di Roma, Piazzale A. Moro 2, 00185 Roma, Italy

Edited by Michael E. Fisher, University of Maryland, College Park, MD, and approved September 17, 2008 (received for review July 18, 2008)

We study the distributions of citations received by a single publication within several disciplines, spanning broad areas of science. We show that the probability that an article is cited c times has large variations between different disciplines, but all distributions are rescaled on a universal curve when the relative indicator $c_f = c/c_0$ is considered, where $c_0$ is the average number of citations per article for the discipline. In addition we show that the same universal behavior occurs when citation distributions of articles published in the same field, but in different years, are compared. These findings provide a strong validation of $c_f$ as an unbiased indicator for citation performance across disciplines and years. Based on this indicator, we introduce a generalization of the h index suitable for comparing scientists working in different fields.

Citation analysis is a bibliometric tool that is becoming increasingly popular to evaluate the performance of different actors in the academic and scientific arena, ranging from individual scholars (1–3), to journals, departments, universities (4), and national institutions (5), up to whole countries (6). The outcome of such analysis often plays a crucial role in deciding which grants are awarded, how applicants for a position are ranked, and even the fate of scientific institutions. It is then crucial that citation analysis is carried out in the most precise and unbiased way.

Citation analysis has a very long history and many potential problems have been identified (7–9), the most critical being that often a citation does not—nor it is intended to—reflect the scientific merit of the cited work (in terms of quality or relevance). Additional sources of bias are, to mention just a few, self-citations, implicit citations, the increase in the total number of citations with time, or the correlation between the number of authors of an article and the number of citations it receives (10).

In this work we consider one of the most relevant factors that may hamper a fair evaluation of scientific performance: field variation. Publications in certain disciplines are typically cited much more or much less than in others. This may happen for several reasons, including uneven number of cited papers per article in different fields or unbalanced cross-discipline citations (11). A paradigmatic example is provided by mathematics: the highest 2006 impact factor (IF) (12) for journals in this category (Journal of the American Mathematical Society) is 2.55, whereas this figure is 10 times larger or more in other disciplines (for example, in 2006, New England Journal of Medicine had IF 51.30, Cell had IF 29.19, and Nature and Science had IF 26.68 and 30.03, respectively).

The existence of this bias is well-known (8, 10, 12) and it is widely recognized that comparing bare citation numbers is inappropriate. Many methods have been proposed to alleviate this problem (13–17). They are based on the general idea of normalizing citation numbers with respect to some properly chosen reference standard. The choice of a suitable reference standard, which can be a journal, all journals in a discipline, or a more complicated set (14), is a delicate issue (18). Many possibilities exist also in the detailed implementation of the standardization procedure. Some methods are based on ranking articles (scientists, research groups) within one field and comparing relative positions across disciplines. In many other cases relative indicators are defined, that is, ratios between the bare number of citations c and some average measure of the citation frequency in the reference standard. A simple example is the Relative Citation Rate of a group of articles (13), defined as the total number of citations they received, divided by the weighted sum of impact factors of the journals where the articles were published. The use of relative indicators is widespread, but empirical studies (19–21) have shown that distributions of article citations are very skewed, even within single disciplines. One may wonder then whether it is appropriate to normalize by the average citation number, which gives only very limited characterization of the whole distribution. We address this issue in this article.

The problem of field variation affects the evaluation of performance at many possible levels of detail: publications, individual scientists, research groups, and institutions. Here, we consider the simplest possible level, the evaluation of citation performance of single publications. When considering individuals or research groups, additional sources of bias (and of arbitrariness) exist that we do not tackle here. As reference standard for an article, we consider the set of all articles published in journals that are classified in the same Journal of Citation Report scientific category of the journal where the publication appears (see details in Methods). We take as normalization the quantity for citations of articles belonging to a given scientific field to be the average number $c_0$ of citations received by all articles in that discipline published in the same year. We perform an empirical analysis of the distribution of citations for publications in various disciplines and we show that the large variability in the number of bare citations c is fully accounted for when $c_f = c/c_0$ is considered. The distribution of this relative performance index is the same for all fields. No matter whether, for instance, Developmental Biology, Nuclear Physics, or Aerospace Engineering are considered, the chance of having a particular value of $c_f$ is the same. Moreover, we show that $c_f$ allows us to properly take into account the differences, within a single discipline, between articles published in different years. This provides a strong validation of the use of $c_f$ as an unbiased relative indicator of scientific impact for comparison across fields and years.

Variability of Citation Statistics in Different Disciplines

First, we show explicitly that the distribution of the number of articles published in some year and cited a certain number of times $c_0$ has a universal behavior when it is rescaled on a universal curve when the relative indicator $c_f = c/c_0$ is considered. Based on this indicator, we introduce a generalization of the h index suitable for comparing scientists working in different fields.

Author contributions: F.R., S.F., and C.C. designed research; F.R., S.F., and C.C. performed research; F.R. analyzed data; and C.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. E-mail: claudio.castellano@roma1.infn.it.

© 2008 by The National Academy of Sciences of the USA
times strongly depends on the discipline considered. In Fig. 1 we plot the normalized distributions of citations to articles that appeared in 1999 in all journals belonging to several different disciplines according to the Journal of Citation Reports classification.

From this figure it is apparent that the chance of a publication being cited strongly depends on the category the article belongs to. For example, a publication with 100 citations is ~50 times more common in Developmental Biology than in Aerospace Engineering. This has obvious implications in the evaluation of outstanding scientific achievements: the simple count of the number of citations received by articles published in a discipline, as measured by the average number of citations per article, can be misleading to assess whether a paper is more successful than one in Aerospace Engineering.

**Distribution of the Relative Indicator $c_f$**

A first step toward properly taking into account field variations is to recognize that the differences in the bare citation distributions are essentially not due to specific discipline-dependent factors, but are instead related to the pattern of citations in the field, as measured by the average number of citations per article $c_0$. It is natural then to try to factor out the bias induced by the difference in the value of $c_0$ by considering a relative indicator, that is, measuring the success of a publication by the ratio $c_f = c/c_0$ between the number of citations received and the average number of citations received by articles published in its field in the same year. Fig. 2 shows that this procedure leads to a very good collapse of all curves for different values of $c_0$ onto a single shape. The distribution of the relative indicator $c_f$ then seems universal for all categories considered and resembles a lognormal distribution. To make these observations more quantitative, we have fitted each curve in Fig. 2 for $c_f \simeq 0.1$ with a lognormal curve

$$F(c) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{[\log(c) - \mu]^2}{2\sigma^2}},$$

where the relation $\sigma^2 = -2\mu$, because the expected value of the variable $c_f$ is 1, reduces the number of fitting parameters to 1. All fitted values of $\sigma^2$, reported in Table 1, are compatible within 2 standard deviations, except for one (Anesthesiology) that is, in any case, within 3 standard deviations of all of the others. Values of $\chi^2$ per degree of freedom, also reported in Table 1, indicate that the fit is good. This allows us to conclude that, in rescaling the distribution of citations for publications in a scientific discipline by their average number, a universal curve is found, independent of the specific discipline. Fitting a single curve for all categories, a lognormal distribution with $\sigma^2 = 1.3$ is found, which is reported in Fig. 2.

Interestingly, a similar universality for the distribution of the relative performance is found, in a totally different context, when the number of votes received by candidates in proportional elections is considered (22). In that case, the scaling curve is also well-fitted by a lognormal with parameter $\sigma^2 \approx 1.1$. For universality in the dynamics of academic research activities, see also ref. 23.

The universal scaling obtained provides a solid grounding for comparison between articles in different fields. To make this even more visually evident, we have ranked all articles belonging to a pool of different disciplines (spanning broad areas of science) according to the unnormalized number of citations to articles published in the same year and in the same field. This is a very natural choice, giving to the numerical value of $c_f$ the direct interpretation as the relative citation performance of the publication. In the literature this quantity is also indicated as the relative performance is found, in a totally different context, when the number of votes received by candidates in proportional elections is considered (22). In that case, the scaling curve is also well-fitted by a lognormal with parameter $\sigma^2 \approx 1.1$. For universality in the dynamics of academic research activities, see also ref. 23.
Table 1. List of all scientific disciplines considered in this article

<table>
<thead>
<tr>
<th>Index</th>
<th>Subject category</th>
<th>Year</th>
<th>(N_p)</th>
<th>(c_0)</th>
<th>(c_{\text{max}})</th>
<th>(\sigma^2)</th>
<th>(\chi^2/\text{df})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Agricultural economics and policy</td>
<td>1999</td>
<td>266</td>
<td>6.88</td>
<td>42</td>
<td>1.0 (1)</td>
<td>0.007</td>
</tr>
<tr>
<td>2</td>
<td>Allergy</td>
<td>1999</td>
<td>1,530</td>
<td>17.39</td>
<td>271</td>
<td>1.4 (2)</td>
<td>0.012</td>
</tr>
<tr>
<td>3</td>
<td>Anesthesiology</td>
<td>1999</td>
<td>3,472</td>
<td>13.25</td>
<td>282</td>
<td>1.8 (2)</td>
<td>0.009</td>
</tr>
<tr>
<td>4</td>
<td>Astronomy and astrophysics</td>
<td>1999</td>
<td>7,399</td>
<td>23.77</td>
<td>1,028</td>
<td>1.1 (1)</td>
<td>0.003</td>
</tr>
<tr>
<td>5</td>
<td>Biology</td>
<td>1999</td>
<td>3,400</td>
<td>14.6</td>
<td>413</td>
<td>1.3 (1)</td>
<td>0.004</td>
</tr>
<tr>
<td>6</td>
<td>Computer science, cybernetics</td>
<td>1999</td>
<td>704</td>
<td>8.49</td>
<td>100</td>
<td>1.3 (1)</td>
<td>0.004</td>
</tr>
<tr>
<td>7</td>
<td>Developmental biology</td>
<td>1999</td>
<td>2,982</td>
<td>38.67</td>
<td>520</td>
<td>1.3 (3)</td>
<td>0.002</td>
</tr>
<tr>
<td>8</td>
<td>Engineering, aerospace</td>
<td>1999</td>
<td>1,070</td>
<td>5.65</td>
<td>95</td>
<td>1.4 (1)</td>
<td>0.003</td>
</tr>
<tr>
<td>9</td>
<td>Hematology</td>
<td>1999</td>
<td>4,423</td>
<td>41.05</td>
<td>1,424</td>
<td>1.5 (1)</td>
<td>0.002</td>
</tr>
<tr>
<td>10</td>
<td>Hematology</td>
<td>1999</td>
<td>6,920</td>
<td>30.61</td>
<td>966</td>
<td>1.3 (1)</td>
<td>0.004</td>
</tr>
<tr>
<td>11</td>
<td>Hematology</td>
<td>2004</td>
<td>8,695</td>
<td>15.66</td>
<td>1,014</td>
<td>1.3 (1)</td>
<td>0.003</td>
</tr>
<tr>
<td>12</td>
<td>Mathematics</td>
<td>2004</td>
<td>8,440</td>
<td>5.97</td>
<td>191</td>
<td>1.3 (4)</td>
<td>0.001</td>
</tr>
<tr>
<td>13</td>
<td>Microbiology</td>
<td>2004</td>
<td>9,761</td>
<td>21.54</td>
<td>803</td>
<td>1.0 (1)</td>
<td>0.005</td>
</tr>
<tr>
<td>14</td>
<td>Neuroimaging</td>
<td>1999</td>
<td>444</td>
<td>25.26</td>
<td>518</td>
<td>1.1 (1)</td>
<td>0.004</td>
</tr>
<tr>
<td>15</td>
<td>Neuroimaging</td>
<td>1999</td>
<td>1,073</td>
<td>23.16</td>
<td>463</td>
<td>1.4 (1)</td>
<td>0.003</td>
</tr>
<tr>
<td>16</td>
<td>Neuroimaging</td>
<td>2004</td>
<td>1,395</td>
<td>12.68</td>
<td>132</td>
<td>1.1 (1)</td>
<td>0.005</td>
</tr>
<tr>
<td>17</td>
<td>Physics, nuclear</td>
<td>1990</td>
<td>3,670</td>
<td>13.75</td>
<td>387</td>
<td>1.4 (1)</td>
<td>0.001</td>
</tr>
<tr>
<td>18</td>
<td>Physics, nuclear</td>
<td>1999</td>
<td>3,965</td>
<td>10.92</td>
<td>434</td>
<td>1.4 (4)</td>
<td>0.001</td>
</tr>
<tr>
<td>19</td>
<td>Physics, nuclear</td>
<td>2004</td>
<td>4,164</td>
<td>6.94</td>
<td>218</td>
<td>1.4 (1)</td>
<td>0.001</td>
</tr>
<tr>
<td>20</td>
<td>Tropical medicine</td>
<td>1999</td>
<td>1,038</td>
<td>12.35</td>
<td>126</td>
<td>1.1 (1)</td>
<td>0.017</td>
</tr>
</tbody>
</table>

For each category we report the total number of articles \(N_p\), the average number of citations \(c_0\), the maximum number of citations \(c_{\text{max}}\), the value of the fitting parameter \(\sigma^2\) in Eq. 1, and the corresponding \(\chi^2\) per degree of freedom (df). Data refer to articles published in journals listed by Journal of Citation Reports under a specific subject category.

“item oriented field normalized citation score” (24), an analogue for a single publication of the popular Centre for Science and Technology Studies, Leiden (CWTS), field-normalized citation score or “crown indicator” (25). In agreement with the findings of ref. 11, \(c_0\) shows very little correlation with the overall size of the field, as measured by the total number of articles.

The previous analysis compares distributions of citations to articles published in a single year, 1999. It is known that different temporal patterns of citations exist, with some articles starting soon to receive citations, whereas others (“sleeping beauties”) go unnoticed for a long time, after which they are recognized as seminal and begin to attract a large number of citations (26, 27).

Other differences exist between disciplines, with noticeable fluctuations in the cited half-life indicator across fields. It is then natural to wonder whether the universality of distributions for articles published in the same year extends longitudinally in time so that the relative indicator allows comparison of articles published in different years. For this reason, in Fig. 4 we compare the plot of \(c_0F(c, c_0)\) vs. \(c_f\) for publications in the same scientific discipline that appeared in 3 different years. The value of \(c_0\) obviously grows as older publications are considered, but the rescaled distribution remains conspicuously the same.

**Fig. 3.** We rank all articles according to the bare number of citations \(c\) and the relative indicator \(c_f\). We then plot the percentage of articles of a particular discipline present in the top 5% of the general ranking, for the rank based on the number of citations (A and C) and based on the relative indicator \(c_f\) (B and D). Different values of \(z\) (different graphs) lead to a very similar pattern of results. The average values and the standard deviations of the bin heights shown are also reported in Table 2. The numbers identify the disciplines as they are indicated in Table 1.
variability is rescaled away if the number of citations (28) is now a well-established standard tool for the evaluation of the scientific performance of scientists. Its popularity is partly due to its simplicity: the h index of an author is the h such that h of his N articles have at least h citations each, and the other N − h articles have, at most, h citations each. Despite its success, as with all other performance metrics, the h index has some shortcomings, as already pointed out by Hirsch himself. One of them is the difficulty in comparing authors in different disciplines.

The identification of the relative indicator $c_f$ as the correct metric to compare articles in different disciplines naturally suggests its use in a generalized version of the h index, taking properly into account different citation patterns across disciplines. However, just ranking articles according to $c_f$, instead of on the basis of the bare citation number, is not enough. A crucial ingredient of the h index is the number of articles published by an author. As Fig. 5 shows, such a quantity also depends on the discipline considered; in some disciplines, the average number of articles published by an author in a year is much larger than in others. However, also in this case, this variability is rescaled away if the number N of publications by a given author in a year is divided by the average value in the discipline $N_0$. Interestingly, the universal curve is fitted reasonably well over almost 2 decades by a power-law behavior $P(N, N_0) \sim (N/N_0)^{-\delta}$ with $\delta = 3.5$ (5).

This universality allows one to define a generalized h index, $h_f$, that factors out also the additional bias due to different publication rates, thus allowing comparisons among scientists working in different fields. To compute the index for an author, his/her articles are ordered according to $c_f = c/c_0$ and this value is plotted versus the reduced rank $r/N_0$ with $r$ being the rank. In analogy with the original definition by Hirsch, the generalized index is then given by the last value of $r/N_0$ such that the corresponding $c_f$ is larger than $r/N_0$. For instance, if an author has published 6 articles with values of $c_f$ equal to 4.1, 2.8, 2.2, 1.6, 0.8, and 0.4, respectively, and the value of $N_0$ in his discipline is 2.0, his $h_f$ index is equal to 1.5. This is because the third best article has $r/N_0 = 1.5 < 2.2 = c_f$, whereas the fourth has $r/N_0 = 2.0 > 1.6 = c_f$.

**Conclusions**

In this article we have presented strong empirical evidence that the widely scattered distributions of citations for publications in different scientific disciplines are rescaled on the same universal curve when the relative indicator $c_f$ is used. We have also seen that the universal curve is remarkably stable over the years. The analysis presented here justifies the use of relative indicators to compare in a fair manner the impact of articles across different disciplines and years. This may have strong and unexpected implications. For instance, Fig. 2 leads to the counterintuitive conclusion that an article in Aerospace Engineering with only 20 citations ($c_f \approx 3.54$) is more successful than an article in Developmental Biology with 100 citations ($c_f \approx 2.58$). We stress that this does not imply that the article with larger $c_f$ is necessarily more “important” than the other. In an evaluation of importance, other field-related factors may play a role: an article with an outstanding value of $c_f$ in a very narrow specialist field may be less important (for science, in general, or for the society) than a publication with smaller $c_f$ in a highly competitive discipline with potential implications in many areas.

Because we consider single publications, the smallest possible entities whose scientific impact can be measured, our results must always be taken into account when tackling other, more complicated tasks, like the evaluation of performance of individuals or research groups. For example, in situations where the simple count of the mean number of citations per publication is deemed to be important, one should compute the average of $c_f$ (not of $c$) to evaluate impact independently of the scientific discipline. For what concerns the assessment of single authors’ performance we have defined a generalized h index (1) that allows a fair comparison across disciplines taking into account also the different publication rates.

Our analysis deals with 2 of the main sources of bias affecting comparisons of publication citations. It would be interesting to tackle, along the same lines, other potential sources of bias, as,

---

**Table 2. Average and standard deviation for the bin heights in Fig. 3**

<table>
<thead>
<tr>
<th>z</th>
<th>$\sigma_z$ (theor)</th>
<th>$z$ (c)</th>
<th>$\sigma_z$ (c)</th>
<th>$z$ (c)</th>
<th>$\sigma_z$ (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.59</td>
<td>4.38</td>
<td>4.73</td>
<td>5.14</td>
<td>0.51</td>
</tr>
<tr>
<td>10</td>
<td>0.81</td>
<td>6.69</td>
<td>7.92</td>
<td>10.07</td>
<td>0.67</td>
</tr>
<tr>
<td>20</td>
<td>1.09</td>
<td>17.68</td>
<td>12.37</td>
<td>20.03</td>
<td>1.15</td>
</tr>
<tr>
<td>40</td>
<td>1.33</td>
<td>35.67</td>
<td>17.48</td>
<td>39.86</td>
<td>2.58</td>
</tr>
</tbody>
</table>

Comparison between the values expected theoretically for unbiased ranking (first 2 columns), those obtained empirically when articles are ranked according to c (3rd and 4th columns), and according to $c_f$ (last 2 columns).
for example, the number of authors, which is known to correlate with a higher number of citations (10). It is natural to define a relative indicator, the number of citations per author. Is this normalization the correct one that leads to a universal distribution, for any number of authors?

Finally, from a more theoretical point of view, an interesting goal for future work is to understand the origin of the universality found and how its precise functional form comes about. An attempt to investigate what mechanisms are relevant for understanding citation distributions is in ref. 29. Further activity in the same direction would definitely be interesting.

**Methods**

Our empirical analysis is based on data from Thomson Scientific’s Web of Science (WOS; www.isiknowledge.com) database, where the number of citations is counted as the total number of times an article appears as a reference of a more recently published article. Scientific journals are divided in 172 categories, from Acoustics to Zoology. Within a single category a list of journals is provided. We consider articles published in each of these journals to be part of the category. Notice that the division in categories is not mutually exclusive: for example, Physical Review D belongs both to the Astronomy and Astrophysics and to the Physics, Particles and Fields categories. For consistency, among all records contained in the database we consider only those classified as “article” and “letter,” thus excluding reviews, editorials, comments, and other published material likely to have an uncommon citation pattern. A list of the categories considered, with the relevant parameters that characterize them, is reported in Table 1. The category Multidisciplinary Sciences does not fit perfectly into the universal picture found for other categories, because the distribution of the number of citations is a convolution of the distributions corresponding to the single disciplines represented in the journals. However, if one focuses only on the 3 most important multidisciplinary journals (Nature, Science, and PNAS), this category fits very well into the global universal picture. Our calculations neglect uncited articles; we have verified, however, that their inclusion just produces a small shift in $c_6$, which does not affect the results of our analysis. In the plots of the citation distributions, data have been grouped in bins of exponentially growing size, so that they are equally spaced along a logarithmic axis. For each bin, we count the number of articles with citation count within the bin and divide by the number of all potential values for the citation count that fall in the bin (i.e., all integers). This holds as well for the distribution of the normalized citation count $c_6$ because the latter is just determined by dividing the citation count by the constant $c_6$, so it is a discrete variable just like the original citation count. The resulting ratios obtained for each bin are finally divided by the total number of articles considered, so that the histograms are normalized to 1.